图书情报工作
第 63 卷 第 18 期 2019 年 9 月

科学数据素养内涵结构研究

■ 秦小燕[1] 初景利[2,3]
[1] 北京航空航天大学图书馆 北京 100083 [2] 中国科学院文献情报中心 北京 100190
[3] 中国科学院大学 北京 100049

摘要：[目的/意义] 对科学数据素养概念的界定与内涵结构的把握，是开展科学数据素养理论研究与教育实践的基础。[方法/过程] 基于信息素养理论、科学数据生命周期理论，在国内外数据素养理论研究与实践进展基础上，界定"科学数据素养"的核心内涵，剖析与之密切相关的各类型素养的源起背景和能力要求，建立科学数据素养的内涵结构体系。[结果/结论] 构建了以科研工作流程和科学数据生命周期为基础的过程结构体系、以新型数据环境下科学数据素养教育目标为核心的目标结构体系，并将数据行为过程与数据培养目标有机联结与统一，构成了科学数据素养过程 - 目标结构体系，明晰了科学数据素养应该达到的目标和具体实现路径，为科学数据素养研究奠定理论基础。

关键词：科学数据素养 内涵 结构
分类号：G252.7 G316
DOI：10.13266/j.issn.0252-3116.2019.18.004

1 引言

素养教育是人类社会发展历程中的永恒话题，不同的历史时期和社会环境，对素养有着不同的定义和要求。当今，我们面对着数据的爆炸式增长和科学技术的不断更新，对数据进行有效识别、获取、处理、评价以及利用的素养就显得尤为重要，直接影响着知识产生、科技创新，乃至日常工作、学习和生活。科学数据素养是数据密集型时代应用而生的新兴概念，对其源起、内涵、发展架构等基本理论问题的学理分析，是开展科学数据素养理论研究与教育实践的基础[1]。本文深入探讨科学数据素养研究的理论基础和科学数据素养的内涵特征，通过剖析其内涵结构，为科学数据素养相关研究提供理论支撑。

2 科学数据素养研究的理论基础

2.1 信息素养理论

"信息素养"亦称"信息素质"，与"数据素养"相比，其历史悠久，概念成熟。它始于 19 世纪末 20 世纪初，80 年代末转化为信息素养。随着信息技术的迅猛发展和信息数量的激增，美国大学和研究图书馆协会(Association of College & Research Libraries, ACRL)于 1989 年发布了关于信息素养的具有里程碑意义的报告，将信息素养确定为信息时代的生存技能，指出"具备信息素养的人，必须能够识别何时需要信息，并能够有效地定位、评估和使用所需信息。"[2]。从此以后，这个定义被广泛使用，人们充分认识到信息素养的重要性，全球许多国家和国际组织为提高信息素养能力而做出大量努力[3]。

信息素养理论最为重要的研究成果，主要包括两方面，其一是关于信息素养内涵结构的讨论，其二是信息素养能力标准与评价体系的形成。关于信息素养内涵结构的研究，有学者提出"三层面信息素养理论"，即信息知识、信息能力和信息文化[4]；也有人认为信息素养由信息意识、信息知识、信息能力、信息道德四方面构成[5]。我国学者皮介郑提出了信息素质的过程结构理论与目标结构理论，并将两种结构理论进行联结，构建了"信息素质过程 - 目标结构体系"[6]，详细解析和描述了信息素质概念的丰富内涵和结构体系，为信息素质教育确立了理论框架。

关于信息素养能力评估，全球很多机构都制定了

作者简介：秦小燕(ORCID:0000-0002-8177-0255)，馆长助理兼文献资源建设部主任，馆员，博士；初景利(ORCID:0000-0003-0723-3985)，中国科学院文献情报中心期刊出版运营总监，中国科学院大学图书情报与档案管理系主任，教授，博士，博士生导师，通讯作者，E-mail: chujl@mail.las.ac.cn。
收稿日期：2019-02-18 修回日期：2019-03-19 本文起止页码：30-39 本文责任编辑：杜杏叶

相应的能力标准与指标体系,其中产生较大影响的包括:美国大学与研究图书馆协会(ACRL)2000 年的“高等教育信息素养能力标准”[7](简称“标准”),英国国立与大学图书馆协会(Society of College, National and University Libraries,SCONUL)2000 年制定及后期修订的信息素养七要素标准[8],澳大利亚与新西兰信息素养研究会(Australian and New Zealand Institute for Information Literacy,ANZIIL)2004 年的“信息素养框架:原则、标准与实务”[9]、以及联合国教科文组织(United Nations Educational, Scientific and Cultural Organization, UNESCO)2013 年发布的“全球媒体与信息素养评估框架”[10]等。以上评估标准尽管各有侧重,但基本上是按照信息处理与利用的过程来建立指标体系,遵循信息需求判断、信息获取、信息评价、信息管理、信息利用等基本流程模式。而 ACRL 于 2015 年颁布的“高等教育信息素养框架”[11](简称“框架”),则进行了重大的改革与调整,面对全球高等教育和学习环境变革、信息生态复杂多变的新形势,改变了 2000 年“标准”中信息素养学习成效或技能列表的线性模式,以相互关联的概念集合来取代,通过一套更丰富的核心理念,可供灵活选择实施,充分挖掘信息素养教育改革的潜力[12]。

“框架”将信息素养延伸至高等教育中学习者的所有学习范畴,并与其他学术社区和社会学习目标相融合。“框架”代表了目前全球信息素养评价的最新理念,创新性地引入“阈概念”和“元素养”这两个核心理念,关注学习者对信息的批判性思维和交互协同能力的元素养,以及通往某一学科领域的门户知识阈概念,在框架指标的表征和描述中,强调一系列动态、灵活的行为方式和批判性、拆解性的思维态度[12]。这些新兴的理念启示我们在理解新的科学研究和高等教育范式中,科学数据素养应该从技能培养到理念提升的变迁。已有的信息素养相关理论成果对于科学数据素养内涵的理解和应用提供了充分的理论基础,也具有实际借鉴意义。

2.2 科学数据生命周期理论

生命周期(Life Cycle),其本义源自生命科学术语,是指一个生物体从诞生、成长、成熟、衰退到死亡所经历的整个过程[13]。这一形象的概念经延伸和扩展后,广泛地应用于自然界和人类社会的其他领域,生命周期理论成为一种重要的研究方法,由于研究对象从产生到消亡的整个过程中处于不同阶段时会呈现各阶段不同的变化及具有各阶段独特的规律,因此,生命周期理论通过分阶段研究来认识整体的研究方法更符合事物发展的一般规律[14]。信息生命周期是将信息视为一种有机体,它具备内在的动态循环规律,从产生、利用到老化消亡,不断进行着自身价值生成、体现和增值的运动过程。科学数据生命周期概念正是借鉴了信息生命周期理论而产生,用于研究科学数据在整个生命周期中从创建到存档、回收的价值衰变过程。科学数据形式多样、载体丰富,其循环过程都有其固有的规律,符合生命周期的研究方法。但科学数据也不完全等同于信息资源“价值老化”的生命周期衰变规律,其生命周期与科研工作流程紧密相关,而且会受到学科属性、研究方法、工具和手段的影响[15]。

目前国外的科学数据理论研究和实践进展迅速,根据地球观测卫星委员会(Committee on Earth Observation Satellites, CEOS)2012 年 4 月发布的调研报告显示,不同类型的科学数据生命周期及管理模型已有 52 个[16],其中几种较为成熟且具有代表性的科学数据生命周期模型包括:

(1)美国政治与社会研究校际联盟(Inter-university Consortium for Political and Social Research, ICPSR)的 DDI(Data Documentation Initiative)组合生命周期模型(Combined Life Cycle Model)[17](如图 1 所示)。主要流程包括 6 个部分:“研究概念→数据收集→数据处理→数据分配→数据发现→数据分析”,其中“数据存档”和“数据再利用”两个环节可因研究需要而成为可选路径,这一模型将数据生命周期与科研流程有效结合起来,不仅使科研人员明确数据与项目之间的关系,也为科研数据服务确立了主线。美国大学图书馆以此模型为基础构建了相应的科研数据服务模型框架,如弗吉尼亚大学图书馆科研生命周期模型[18]、密歇根州立大学记录生命周期模型[19]等。

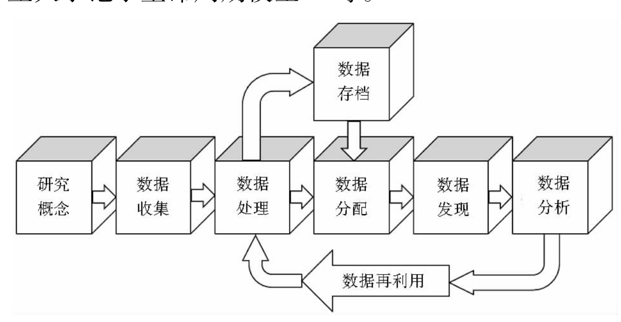


图 1 DDI 组合生命周期模型[17]

(2)英国数据保存中心(Digital Curation Center, DCC)针对数字对象或数据集提出了“数据监管生命周期模型”,如图 2 所示。以数据为核心,外围由环状层次结构组成,第一层是数据描述与表征,第二层是数据

长期保存计划，第三层是数据利益相关者的数据分享与活动参与，第四层是数据管理与长期保存，第五层是与数据监管相关的全生命周期各阶段："创建或接收数据→评估和选择→数据获取→长期保存行动→数据存储→数据获取与再利用→数据转换"[20]。

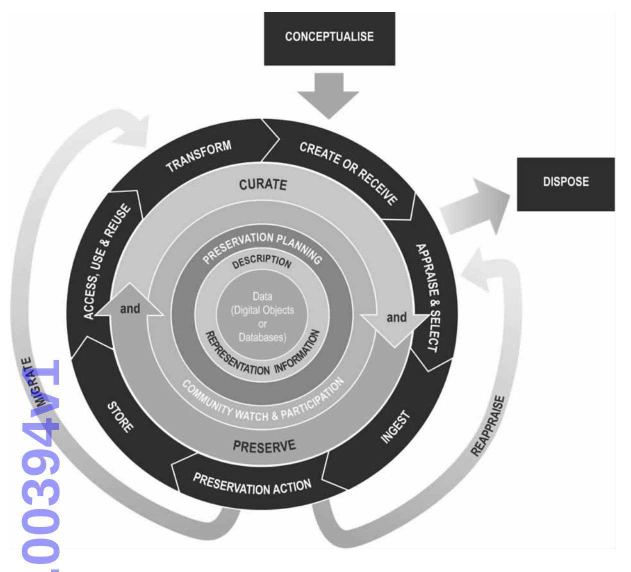


图 2　DCC 数据监管生命周期模型

（3）英国数据存档（UK Data Archive，UKDA）科学数据生命周期，数据经历以 6 个阶段为主的循环往复结构："创建 – 处理 – 分析 – 保存 – 获取 – 再利用"，其寿命比科研项目更长[21]。

（4）美国自然科学基金（National Science Foundation，NSF）2009 年启动的 Data ONE（Data Observation Network for Earth）科学数据生命周期模型，归纳了 8 个主要阶段："计划 – 收集 – 确认 – 描述 – 保存 – 发现 – 整合 – 分析"，是一个循环结构[22]。

综上，目前的科学数据生命周期模型，基本上是围绕科学数据整个生命过程所经历的各个阶段来展开的，有研究认为科学数据生命周期是一种线性结构，更多的研究认为是循环结构，由于科学数据是科学研究的产物，因此与科研项目的研究流程密不可分。虽然自然科学与人文社会科学的数据生命周期，在数据迁移、数据转换、数据生命周期终点等方面存在一定差异，但这并不影响科学数据的某些固有属性，尽管学科不同，但在模型中存在一些相同或相似阶段，如数据发现、数据收集、数据处理、数据分析等，这启发我们在研究科学数据素养的内涵结构时，可以从科学数据的固有属性出发，构建学科通用型科学数据素养内涵体系，在此基础上进行补充并扩展具有学科特性、研究方法特点的学科指导框架。

3　科学数据素养的内涵与特征

与许多新兴和跨学科领域一样，由于时代背景、学术传承与研究视角的差异，数据素养对于不同的利益相关者和专家学者意味着不同的内涵和意义。对于不同研究情境中的数据素养内涵，并不能做出明确的界定。"科学数据素养"，既属于"数据素养"的研究领域，又有其特定的概念范畴，因此，有必要对"科学数据素养"的内涵与特征进行具体的阐述与剖析。

3.1　科学数据的内涵

3.1.1　数据　传统意义上的"数据"（Data）是指"有根据的数字"。在信息化、网络化的双重驱动下，数据的内涵和外延不断扩大，它不仅指代"数字"，还统称一切存储在计算机上的信息，包括文本、声音、视频等。美国国家自然科学基金（NSF）在 2005 年发布的"长期保存的数字数据集"中指出，数据指的是任何可以以数字形式存储的信息，包括文本、数字、图片、视频或影片、音频、软件、算法、公式、动画制作、模型、仿真等[23]。我国国标（GB/T 18391.1 – 2009）中将"数据"定义为："对事实、概念或指令的一种形式化表示，适用于以人工或自动方式进行通信、解释或处理"[24]。数据可以代表很多不同的事物，并且有许多不同的分类方法，按照其产生方式，数据可以分为观测型、实验型、仿真型、派生或编译型、引用或规范型[25]。

随着大数据、云计算、移动互联网、物联网、人工智能的发展，从数字化和计算机角度出发，可以把所有的内容、媒介或出版物都看成是数据集合，所有这些数据对象都可以数字化、结构化、语义化地解析、标注、链接，成为有逻辑意义和相互关联的知识对象；所有这些知识对象都可计算、可重组、可融汇、可再创造，而且可以和使用者进行个性化、动态化的交互以形成新的数据对象和知识内容。各类数据对象成为显性、基础、首要的知识对象，当前我们正从信息环境过渡到数据环境[26]。

当今时代，数据的概念已经远远超出已有的数据定义所表述的意思，由于其形式多样、内涵丰富，实质上数据已经演化为信息、情报、知识和语义的统称。学术交流体系也因此发生了重大变化，数据作为研究的重要基础资源和科研成果产出形式，在科学研究过程中发挥着重要的作用。由于数据类型极其丰富，不同领域，有着不同的侧重点和理解，包括政府数据、医疗数据、商业数据等等，本文重点研究的是科学数据。

3.1.2　科学数据　什么是科学数据？国外的主要名

称表述包括“Science Data(科学数据)”“Scientific Data(科学数据)”或“Research Data(研究数据)”。国内则经由“科技数据”演化为“科学数据”[27]。关于科学数据的定义，国内外研究机构或学者主要有以下一些观点：

(1)传统意义的科学数据。我国科学数据共享工程中对科学数据的定义如下：是指人类在认识世界、改造世界的科技活动中所产生的原始性、基础性数据，以及按照不同需求系统加工的数据产品和相关信息。它既包括了社会公益性事业部门所开展的大规模观测、探测、调查、实验和综合分析所获得的长期积累与整编的海量数据，也包括国家科技计划项目实施与科技工作者长年累月科学实践所产生的大量数据[28]。经济合作与发展组织[29]、美国国立卫生研究院[30]、澳大利亚国立大学[31]、剑桥大学[32]等机构都从相关角度对“研究数据(Research Data)”做了明确定义。对于专业领域的科学数据，许多机构或组织也对其进行了相关阐释，如我国农业科学数据共享中心将农业科学数据定义为：从事农业科技活动所产生的基本数据，以及按照不同需求而系统加工整理的数据产品和相关信息[33]。

综合以上的研究成果，各机构从科学数据管理的角度出发，对科学数据的概念进行了相关界定，大致可以分为3类：一类是指那些能对研究成果进行验证的实际记录材料，而不包括科学研究过程中产生的笔记、初步的实验分析结果等未经过验证的数据、资料；第二类是将与科学研究相关的实验数据、实验笔记、图像、音视频、模拟系统等半成品作为科学数据范畴；第三类则是含义最为广泛的理解，即包括科学研究全过程所产生的过程数据、半成品以及研究成果等。

(2)数字化科研背景下的科学数据。随着信息技术的进步，科学研究的信息化日益成熟，科学计算促使人类描述社会复杂事物的能力不断提升。各种网络传输系统和数据存储与分析设施不仅有助于科学家获得强大的观察能力、分析能力、甚至实验能力，还能促进科研方法从理论分析和观察向科学研究对象的模拟与仿真发展，并推动“以数据为基础的科学研究第四范式”的形成，由假设驱动向直接基于科学数据进行探索的科学方法转变[34]。科学研究和发现范式的转变给科学数据带来了新的含义，具体包括以下几个方面：

首先，通过探测器等高端设备的采集、高性能计算机模拟等产生大量的、原始的科学数据。这些科学数据不仅包括从传统条件下的理论预测和实验观测所获得的实验观测结果，还包括研究人员、科学仪器甚至科学研究过程和管理机制等科学研究活动各个方面的因素综合，以及通过计算机仿真和模拟分析等方式产生的数字表达等，也包括社会科学领域的研究对象。这种数字表达优于传统表达的特点在于可以描述大规模或微尺度实体，根据科学研究所需，进行各种形式的组合、变化和数字表达[15]。

其次，初始数据包括科学实验未经处理的数据及科学研究对象的数字表达。从初始数据到中间数据乃至最终研究结果的科学工作流技术，使得科学数据的可视化方法、技术及软件资源也以一种资源的形式存在，人们能以一种可重复、可验证、分布式的方式来描述科学研究或科学实验过程，了解科学研究过程中所采用的数据处理方法、模型和工具，从而对科学研究进行验证[35]。

另外，大数据时代的“数据”，不再是孤立的、静态的数据，而是动态的、系统化的数据，以持续且不间断的“数据流”形式而存在，成片的、互为关联、有生命力的数据。数据成为社会各行各业创新必不可少的基础资源，对其进行系统采集、科学分析至关重要。

(3)本文研究的科学数据。鉴于传统科研环境、E-research和大数据时代的发展，科学数据的产生途径、组织形式以及应用领域发生着巨大变化。本文所研究的科学数据，指的是在自然科学、工程技术科学、人文社会科学等领域，人类在认识世界、改造世界的各类科技活动中，通过观察、实验、调查记录或计算等不同方式所产生的适合于通信、解释或处理的原始数字化信息，以及按照不同科研需求进行系统加工或重组的数据产品和相关信息，包括数值型数据、文本数据、图像、音频、视频等多媒体数据等，且可用于科学研究、技术设计、查证、决策等。

3.2 科学数据素养的内涵

“数据素养”较早源起于美国教育界，随着科技发展，数据素养的内涵也发生着一定的演变，在不同时代和不同领域中存在差异，目前已有的国内外研究中尚未对数据素养的概念形成标准性的定论。在术语上主要有以下几种表述方式：“数据素养”(Data Literacy)、“数据信息素养”(Data information literacy)、“科学数据素养”(Science data literacy)、“研究数据素养”(Research data literacy)、“数据管理素养”(Data management literacy)等。

其中，“科学数据素养”的提出，将“数据素养”的概念和应用范畴限定于科学研究领域，而且对“数据”

的范围给予了更加明确的界定,特指"科学数据",即人类在认识世界、改造世界的科技活动中所产生的原始性、基础性数据,以及按照不同需求系统加工的数据产品和相关信息。素养,被定义为"读写能力"或"由训练或实践而获得的一种修养",包括道德品质、外表形象、知识水平与能力等各个方面。从这个定义出发,"科学数据素养"可被理解为"在科学研究过程中,科研人员作为数据的生产者、使用者和管理者,收集、处理、管理、评价和利用数据进行科学研究所涉及的思维、知识和技能,以及在数据生命周期中普遍遵循的伦理道德与行为规范。"

科学数据素养,强调对科学数据的理解、利用和管理能力,目的是将数据转化为知识。信息技术和数字信息资源的快速发展使科学研究活动发生巨大变革,一方面,新的科研环境对研究人员的数据利用能力提出了新的要求,另一方面,数据密集型科研范式对数据管理能力及其相关专业人员的需求显著增加。研究人员不仅需要对科学数据进行有效地分析与利用,还需要具备对科学数据进行系统收集和管理的能力。纵观现有的科学数据素养研究,基本围绕两条路径展开,一是数据利用视角,关注信息素养视域下的用户数据行为研究,即如何利用与再利用科学数据;二是数据管理视角,即关注科研生命周期的数据生产、组织和存储等环节。两个研究目标既有区别又相互关联,其根本是培养学习者在科学探索中收集、加工、管理、评估和使用数据的知识和技能,使其具备适应科研创新与发展所需要的科学数据素养能力。

3.3 科学数据素养与其他素养的关系

"素养"是一个随时代变迁而不断发展和丰富的概念,当一种行为或生活方式变得越来越大众化而且有影响力时,传统素养的作用和价值日渐削弱,客观上需要倡导并提出一种新的素养,使个人能够发挥自己的知识和潜力,并充分参与其社区以及更广泛的社会活动,来适应社会发展带来的挑战和要求,实现自己的目标。可以说"科学数据素养"就是在与之相关的一些素养概念基础上产生的,诸如信息素养(Information Literacy)、数字素养(Numeric Literacy)、科学素养(Science Literacy)、统计素养(Statistical Literacy)等,厘清科学数据素养与这些相关素养的联系与区别是理解科学数据素养的前提。

3.3.1 与信息素养的关系 信息素养概念自 1974 年诞生以来,引起了世界范围内的研究与讨论,最具影响力的是 ACRL 在 2000 年发布的《高等教育信息素养能力标准》,引用了美国图书馆协会(American Library Association,ALA)在 1989 年给出的比较权威的定义,认为信息素养是指"个人能认识到何时需要信息,有效地搜索、评估和使用所需信息的能力"[7]。时至今日,人类的信息技术手段与信息环境发生了翻天覆地的改变,信息素养的内涵也被从多侧面、多角度进行了阐释。ACRL 在 2015 年发布的《高等教育信息素养框架》,将信息素养概念丰富为"包括反思性信息发现,理解信息如何产生与评估,以及利用信息创造新知识、合理参与学术社团的一组综合能力"[11],强调动态性、灵活性、与各种新兴素养的联系与融合等。

近年来,随着 E-science 和 E-social science 的深入发展以及数据密集型科研范式的兴起,在科学研究中,越来越多的研究建立在对已有数据的重新组织和分析利用基础之上,科学数据的重要性和价值尤其凸显。信息素养的内涵和外延也随着大数据时代的到来而发生改变,在新的信息生态环境中,以科学数据管理和利用为核心的科学数据素养,应该成为当前高等教育中信息素养的重要内容之一。在某种程度上,可以将科学数据素养看作是信息素养的延伸和扩展,正如相关研究机构及专家学者所指出,应该把研究数据看成一种信息,将数据管理知识和技能纳入信息素养定义中,确保学生获得职业生涯成功的必备技能[36-39]。然而,在实际应用过程中,科学数据素养与传统的信息素养有着本质区别,这种差异的根源主要在于处理数据的复杂性比任何其他类型的信息都要大。信息素养反映的是查找、获取、评价和利用信息的能力,而科学数据素养反映的是理解、使用、管理科学数据的能力,以关注数据的特殊属性及应用价值为核心。

3.3.2 与数字素养的关系 "数字素养"是 P. Gilster 于 1997 年正式提出,并将其定义为"理解、评估和使用基于计算机技术的各种数字资源和信息的能力"[40],并将批判性思考能力作为数字素养的核心,而不仅仅是数字技术利用技能。以色列学者 Y. E. Alkalai 根据多年研究和工作经验,构建了数字素养概念的 5 个框架,包括图片-图像素养、再创造素养、分支素养、信息素养和社会-情感素养,其含义分别为"识别理解图形图像信息的能力""重新整合信息的能力""非线性的信息搜索""在零散的信息中建构知识的能力""检索、筛选、辨别、使用信息的能力""共享知识,进行数字化情感交流的能力"[41]。国际图联 2017 年发布的数字素养宣言中,认为"具备数字素养意味着可以在高效、有效、合理的情况下最大限度地利用数字技术,以满足

秦小燕，初景利. 科学数据素养内涵结构研究[J]. 图书情报工作,2019,63(18):30－39.

个人、社会和专业领域的信息需求"[42]。可见,数字素养强调掌握和应用数字技术与网络知识,确定、组织、理解、评估和创造信息,以满足学习、工作、生活等各方面的需要,是一个终身学习的过程。在数据时代,当我们强调作为数据生产者、使用者和管理者,所应具备的科学数据素养能力时,必然无法脱离数字素养的基本要求,但鉴于数据的丰富类型及其复杂的处理流程,科学数据素养的研究范畴将更加广泛而且专深。

3.3.3 与科学素养的关系 通过对"科学素养"概念的考察,国际上普遍将其概括为3个组成部分,即了解科学知识、掌握科学的研究过程和方法、了解科学技术对社会和个人所产生的影响[43－44],科学素养是研究者在科学研究过程中所表现出来的基本素质和修养,是科研创新与科研任务顺利完成的基本保障,这意味着每个从事科学研究的人都应该具备科学素养,它涉及整个科学思维体系、价值观、文化认同和科学信息交流模式,直接关系到科研成果的阅读、理解与产出[45]。科学数据素养与科学素养密切相关,都涉及与科学思考和科学研究相关的方法、途径、态度和技能,科学数据素养丰富了科学素养的内涵。

3.3.4 与统计素养的关系 "统计素养"是指理解并合理使用统计数据和统计方法以解决实际问题的能力,并在了解多方面信息的基础上做出决策[46]。全国义务教育数学课程标准中对统计素养有一个较为全面的解释:"能从统计的角度思考与数据信息有关的问题;能通过收集数据、描述数据、分析数据的过程做出合理的决策,认识到统计对决策的作用;能对数据的来源、处理数据的方法,以及由此得到的结果进行合理的质疑"[47]。由此可见,统计素养的核心是对数据的辩证思考与应用,计算能力也是其中一个重要因素。在大数据时代,统计素养,作为理解和利用包括方法在内的统计信息的能力,与数据科学有着较强的相关性,数据科学和统计学之间的传统界限已经变得越来越模糊了。因此,有学者提出随着数据的作用和性质发生改变,统计素养的传统定义应该得到更新,以体现数据在我们生活中所发挥的极大的作用,认为数据素养为统计素养赋予了更为丰富的内涵[48]。

通过对新信息环境下相关素养的深入剖析,包括信息素养、数字素养、科学素养、统计素养,分析其各自的内涵特征与内在联系,可以更好地把握科学数据素养的核心内涵与研究范畴。笔者认为,具备科学数据素养的人应该同时具备信息素养、数字素养、科学素养和统计素养的相关方面,其核心是将批判性思维贯穿于数据利用与处理的全过程,有效利用数据解决科学问题的能力。

4 科学数据素养的内涵结构

科学数据素养到底应该包含哪些具体的内涵要素,这些要素有着怎样的内在联系与结构,如何才能系统而明确地描绘出科学数据素养的内涵结构呢?从以上对科学数据素养内涵的讨论中可以认识到,科学数据素养的内涵是相当丰富的,对其结构的准确把握,是确定科学数据素养教育理论框架的基本前提。

4.1 科学数据素养的过程结构

为了准确而完整地把握科学数据素养的内涵及结构,通过考察大量国外的科学数据素养项目、教育实践模式及课程大纲,发现对科学数据素养概念的运用,不同表述之间有着内在的相似性,一般都是在明确了科学数据素养的定义之后,按科学数据的利用行为或数据问题的解决过程来展开的,这个过程无论从内容还是秩序上都保持高度的稳定性,是由多个逻辑环节共同组成的,每个环节对数据行为主体在认知、技能、情感等方面的要求,就构成了科学数据素养的内涵。结合科研工作流程和科学数据生命周期,本文构建了科学数据素养的过程结构体系,见图3。

从整个过程结构体系可以看出,在数字化科研以及大数据时代,以科学数据为核心的行为过程可以凝练为以下几个阶段:数据需求分析、数据生产与收集、数据分析与处理、数据出版与共享、数据组织与保存、数据发现与获取、数据评价与再利用。整个过程是纵向发展的,而且在过程中的每一个阶段都会对行为主体提出相应的知识和技能要求,以及对意识、观念和伦理道德方面的具体要求,如关于数据源、数据环境的知识,对数据处理技术的掌握,数据获取与评价的能力,以及整个过程中应该遵循的道德和规范等。同时,行为主体的态度和情感也在发生作用,影响着整个过程的进行。

4.2 科学数据素养的目标结构

通过对相关文献的调研,除了从数据行为过程本身出发来理解科学数据素养内涵并进行相关研究之外,还有部分研究,尤以国内学者为主,他们对科学数据素养内涵的认识,以及对其概念的应用,通过高度概括和抽象的表述,提炼出科学数据素养的教育目标,即在新的大数据环境下,受教育者应该达到哪几方面的目标要求即可被认为具备科学数据素养[49－52]。对于科学数据素养内涵和目标的探讨,虽然不同的研究者

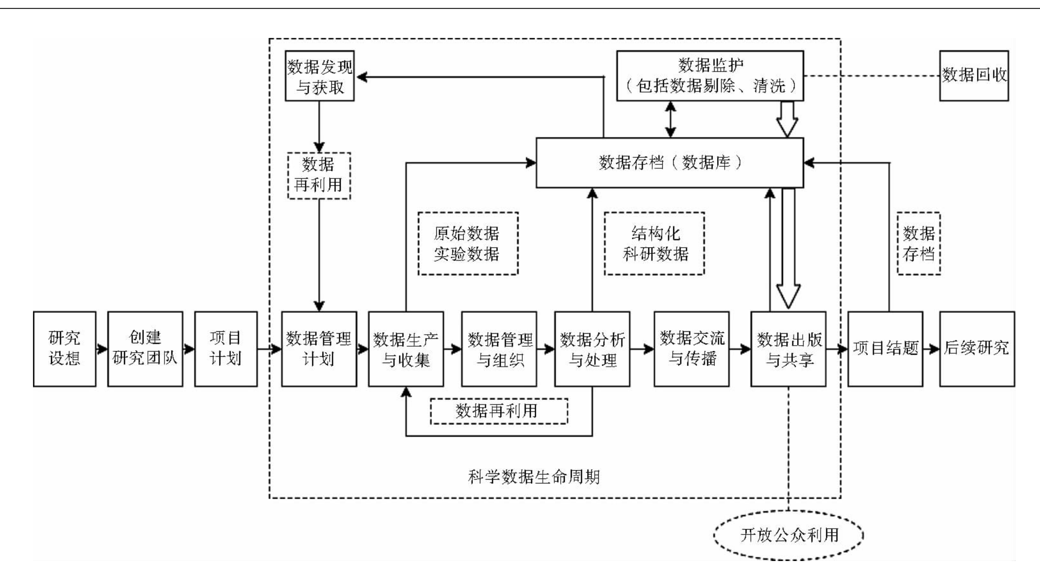


图3　科学数据素养过程结构体系

有着自己对各组成部分概念的理解和外延运用的表述形式，但本质上存在一些共性特征，即都包含了数据知识、数据技能和数据认知3方面的内容，并形成以数据技能为中枢，联结数据知识和数据认知的3层结构。其中，数据知识是基础，围绕数据的所有活动在此基础之上展开，伴随数据知识的积累和应用即形成数据技能，数据技能的发展不仅可以丰富数据知识，还可以加深人们对数据的认知。数据认知主要包括两方面的内容，一方面是人们对数据价值的认识和研究兴趣，即数据意识和数据文化，另一方面是认识到数据发展中的各种行为规范和道德准则，即数据伦理，保证数据主体的数据行为遵循正确的方向，从而维护数据生态的正常秩序，确保人们在数据活动中的利益和社会整体利益的一致性。经过对主要观点的整理和凝练，可以提取出5个相辅相成、互为基础、紧密关联的部分，共同构成科学数据素养概念的完整内涵，并具有相对稳定的理解和表述，分别为：数据意识、数据文化、数据知识、数据技能、数据伦理。

这5部分内容也应成为我们设计和实施科学数据素养能力指标体系和教育活动时所追求的目标内涵，这种内涵结构理论，本文称之为科学数据素养目标结构体系，见图4。

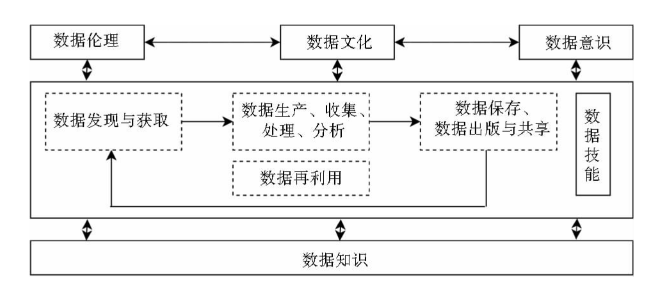


图4　科学数据素养目标结构体系

4.3　科学数据素养的过程－目标结构

从前面的叙述来看，科学数据素养本身包含了信息素养、数字素养、科学素养和统计素养等多方面素养的共同之处，另外也有自己独特的可区别于这些素养的特点。由于科学数据素养的复杂性，对其概念内涵的理解存在难度，因此，能否准确理解这个概念内涵将关系到能否准确建立科学数据素养的评价标准、教育实践活动内容。同时，能否准确把握科学数据素养的内涵也是制定科学数据素养能力评价体系的必要条件。本文对科学数据素养内涵的研究，深入剖析了国内外学者对此类相关素养内涵研究相同的出发点，即围绕数据生态中的关键性角色进行：①个体。个体的数据行为过程有着特定起源和发展历史，这是科学数据素养概念本身应该具有的内涵结构；②社会。科学数据素养教育应该达到的目标是客观社会现实的要求。综观国内外学者对相关素养内涵研究采用的方法和最终构建的内涵结构，虽然他们在思路方向和表述方式中存在一定差异，但本质上具备很多可以将二者联系和统一起来的共性特征。在科学数据素养的内涵研究中，也可以将这些共性特征联系统一后形成科学数据素养过程－目标结构体系，如图5所示，对过程中每个阶段的要求可归结为目标，目标的具体表现形式则是行为过程，目标经认知、行为和情感层面的升华和内化最终形成科学数据素养。

科学数据素养过程－目标结构体系的建立，更加全面、完整地阐释了科学数据素养的内涵及结构，体现了科学数据素养的本质意义与实际价值。能够提高科学数据素养教育实践的可操作性，也有利于相关人员

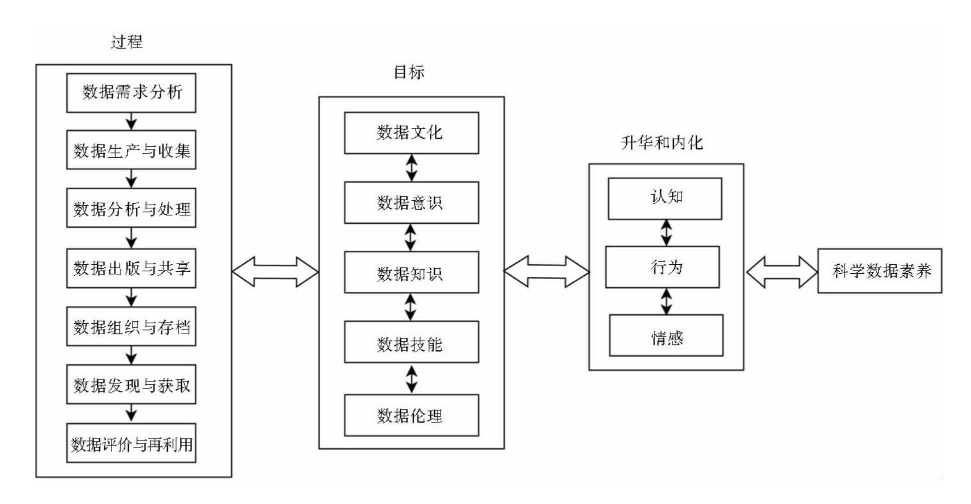


图5 科学数据素养过程-目标结构体系

在科学数据素养实践中进行考核评价和运用实施。这是一个动态的理论分析框架,具体内容将随着数据时代的发展得到不断地自我调整和自我发展,从而形成科学、合理、动态、稳定的科学数据素养体系。

5 结语

对科学数据素养概念的界定与内涵结构的把握,是研究科学数据素养教育理论框架的基础。本文基于信息素养理论、科学数据生命周期理论,在国内外数据素养理论研究与实践进展基础上,重点界定了"科学数据"的具体范畴,以及由此而确定的"科学数据素养"的内涵与特征。通过对信息素养、数字素养、科学素养、统计素养这些与科学数据素养密切相关的各类型素养的源起背景和能力要求的剖析,加深了对科学数据素养的核心内涵与研究范畴的理解。

构建了科学数据素养内涵结构体系,包括以科研工作流程和科学数据生命周期为基础的过程结构体系、以新型数据环境下科学数据素养教育目标为核心的目标结构体系,并将数据行为过程与数据培养目标有机联结与统一,构成了科学数据素养过程-目标结构体系,明晰了科学数据素养应该达到的目标和具体实现路径,为科学数据素养相关研究奠定了理论基础。

参考文献:

[1] 秦小燕.科学数据素养能力指标体系构建与实证研究[D].北京:中国科学院大学,2018.

[2] Association of college and research libraries. Presidential committee on information literacy: final report[R/OL].1989.[2018-01-10]. http://www.ala.org/acrl/publications/whitepapers/presidential.

[3] HORTON F W. Understanding information literacy: a primer. Information society division, communication and information sector UNESCO, Paris[EB/OL].[2018-01-10]. http://unesdoc.unesco.org/images/0015/001570/157020e.pdf.

[4] 秦殿启,张玉玮.三层面信息素养理论的建构与实践[J].情报理论与实践,2017,40(6):13-17.

[5] 韩静娴,赵曼娟.信息素养教育理论与实践[M].广州:世界图书广东出版公司.2014.10:1.

[6] 皮介郑.信息素质理论与教育研究[D].北京:中国科学院研究生院(文献情报中心),2003.

[7] Association of college and research libraries. Information literacy competency standards for higher education[R/OL].[2018-02-01]. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/standards.pdf.

[8] Society of college, national and university libraries. The seven pillars of information literacy[EB/OL].[2018-02-01] https://www.sconul.ac.uk/page/seven-pillars-of-information-literacy.

[9] BUNDY A. Australian and New Zealand information literacy framework: principles, standards and practice[M]. 2nd ed. Adelaide, SA, Australia: Australian and New Zealand Institute for Information Literacy. 2004.

[10] UNESCO. Global media and information literacy assessment framework: country readiness and competencies. [EB/OL]. [2018-08-10]. http://unesdoc.unesco.org/images/0022/002246/224655e.pdf.

[11] Association of college and research libraries. Framework for information literacy for higher education[R/OL]. [2018-08-10]. http://www.ala.org/acrl/standards/ilframework.

[12] 秦小燕.美国高校信息素养标准的改进与启示——ACRL《高等教育信息素养框架》解读[J].图书情报工作,2015,59(19):139-144.

[13] 朱晓峰.生命周期方法论[J].科学学研究,2004(6):566-571.

[14] 马费成,夏永红.网络信息的生命周期实证研究[J].情报理论与实践,2009,32(6):1-7.

[15] 刘霞.高校科学数据管理与实证研究[M].武汉:武汉理工大学出版社.2013.09.

[16] CEOS. Data life cycle models and concepts[EB/OL]. [2018-02-10]. http://wgiss.ceos.org/dsig/whitepapers/.
[17] DDI structural reform group. Overview of the DDI version 3.0 conceptual model[EB/OL]. [2018-02-12]. http://opendatafoundation.org/ddi/srg/Papers/DDIModel_v_4.pdf.
[18] University of Virginia library. Research data services + sciences [EB/OL]. [2018-11-06]. http://data.library.virginia.edu/data-management/.
[19] Michigan State University. Records management[EB/OL]. [2018-02-12]. http://archives.msu.edu/records/.
[20] HIGGINS S. The DCC curation lifecycle model[J]. International journal of digital curation, 2008, 3(1): 134-140.
[21] UK data archive. Research data lifecycle[EB/OL]. [2018-02-12]. http://www.data-archive.ac.uk/create-manage/life-cycle.
[22] MICHENER W K, JONES M B. Ecoinformatics: supporting ecology as a data-intensive science[J]. Trends in ecology & evolution, 2012, 27(2): 85-93.
[23] NSF. National science board. Long-lived digital data collections: Enabling research and education in the 21st century. Washington, D. C.: National Science Board, National Science Foundation. [EB/OL]. [2018-03-26]. http://www.nsf.gov/plus/2005/nsb0540/nsb0540.pdf.
[24] 中华人民共和国国家质量监督检验检疫总局. 信息技术 元数据注册系统(MDR) 第1部分:框架. GB/T18391.1-2009[S]. 北京:中国标准出版社,2009.
[25] University of Virginia library. Data types & file formats[EB/OL]. [2018-03-26]. http://data.library.virginia.edu/data-management/plan/format-types/.
[26] 张晓林. 颠覆性变革与后图书馆时代——推动知识服务的供给侧结构性改革[J]. 中国图书馆学报,2018(1):1-12.
[27] 孙九林,黄鼎成,李晓波. 我国科技数据管理和共享服务的新进展[J]. 世界科技研究与发展,2002(5):15-19.
[28] 科学数据共享工程技术标准 SDS/T1003.2-2004. 科学数据共享概念与术语[S/OL]. [2018-03-11]. http://www.sciencedata.cn/pdf/2.pdf.
[29] OECD. OECD principles and guidelines for access to research data from public funding[R/OL]. [2018-05-16]. http://www.oecd.org/dataoecd/9/61/38500813.pdf.
[30] National institutes of health. NIH grants policy statement. [EB/OL]. [2018-03-17]. http://grants.nih.gov/grants/policy/nihgps_2011/nihgps_ch2.htm.
[31] The Australian national university. Information literacy program. ANU data management manual: managing digital research data at the Australian national university[EB/OL]. [2018-01-20]. https://services.anu.edu.au/files/DataManagment.pdf.
[32] University of Cambridge. Explanation of terms[EB/OL]. [2018-03-27]. http://www.lib.cam.ac.uk/preservation/incremental/glossary.html.
[33] 农业科学数据共享中心简介[EB/OL]. [2018-01-15]. http://www.agridata.cn/homepage/ch_intro.asp.
[34] HEY T, TANDLEY S, TOLLE K. 第四范式:数据密集型科学发现[M]. 潘教峰,张晓林,译. 北京:科学出版社,2012.
[35] 钱鹏. 高校科学数据管理研究[D]. 南京,南京大学,2012.
[36] ANDRETTA, S, POPE, A, WALTON, G. Information literacy education in the UK[J]. Communications in information literacy, 2008,2(1):36-51.
[37] Research information network. The role of research supervisors in information literacy[R/OL]. [2018-10-22]. http://www.rin.ac.uk/system/files/attachments/Summary_of_surveys.pdf.
[38] SCHNEIDER R. Research data literacy [C]. European conference on information literacy. Springer international publishing, 2013, 397: 134-140.
[39] Vitae. Researcher Development framework[R/OL]. [2018-05-08]. https://www.tees.ac.uk/docs/DocRepo/Research/Vitae-Researcher-Development-Framework.pdf. [40] GILSTER P. Digital literacy[M]. New York: Wiley,1997:25-28.
[41] ESHET-ALKALAI Y. Digital literacy: a conceptual framework for survival skills in the digital era[J]. Journal of educational multimedia and hypermedia,2004,13(1):93-106.
[42] 何蕾. 国际图联数字素养宣言[J]. 图书馆论坛,2017,37(11):1-4.
[43] SHAMOS M H. The myth of scientific literacy[M]. New Brunswick, NJ: Rutgers University Press. 1995.
[44] JENKINS E W. Scientific literacy: functional construct[M]. BAKER D, CLAYJ, FOX C. In Challenging ways of knowing in English, maths, and science, London: Falmer Press, 1996:43-51.
[45] ELMBORG J. Critical information literacy: implications for instructional practice[J]. Journal of academic librarianship 2006(32):192-199.
[46] 李俊. 论统计素养的培养[J]. 浙江教育学院学报,2009(01):10-16.
[47] 中华人民共和国教育部. 全日制义务教育数据课程标准:实验稿[M]. 北京:北京师范大学出版社,2001.
[48] GOULD R. Data literacy is statistical literacy[J]. Statistics education research journal 2017,16(1):22-25.
[49] 张晓阳,李楣. 基于胜任特征的研究生数据素养能力测评量表研究[J]. 图书情报工作,2017,61(8):89-95.
[50] 张斌,刘三女牙,刘智,等. 基于大数据的师范生数据素养培养策略研究[J]. 电化教育研究,2017,38(12):86-91,120.
[51] 凌婉阳. 大数据与数据密集型科研范式下的科研人员数据素养研究[J]. 图书馆,2018(1):81-87.
[52] 卜冰华. 大数据环境下我国大学生数据素养教育研究综述[J]. 数字图书馆论坛,2017(12):63-67.

作者贡献说明:

秦小燕:资料调研与论文撰写;

初景利:修改、审定论文。

Research on the Connotation Structure of Scientific Data Literacy

Qin Xiaoyan[1] Chu Jingli[2,3]
[1] Library of Beihang University, Beijing 100083
[2] National Science Library, Chinese Academy of Sciences, Beijing 100190
[3] University of Chinese Academy of Sciences, Beijing 100049

Abstract: [**Purpose/significance**] The definition of the concept of scientific data literacy and the grasp of its connotation structure are the basis for the theoretical research and educational practice of scientific data literacy. [**Method/process**] Based on the information literacy theory and the scientific data life cycle theory, this paper defined the core connotation of "scientific data literacy" on the basis of theoretical research and practical progress at home and abroad, analyzed the origin background and ability requirements of various types of literacy closely related to it, and established the connotation structure system of scientific data literacy. [**Result/conclusion**] This paper constructed the process structure system based on the scientific research workflow and scientific data life cycle, and the object structure system centered on scientific data literacy education goals under the new data environment. In addition, the data behavior processes and data training objectives were connected and unified organically, and the Process-Object structure system of the scientific data literacy was constituted, which clarifies the goals and specific implementation paths that scientific data literacy should achieve. Finally, the above work will lay a theoretical foundation for scientific data literacy research.

Keywords: scientific data literacy connotation structure